RR-89-51-ONR

(2)

AD-A215 504

INTEGRATING COGNITIVE AND PSYCHOMETRIC MODELS TO MEASURE DOCUMENT LITERACY

Kathleen Sheehan
and
Robert J. Mislevy

This research was sponsored in part by the
Cognitive Science Program
Cognitive and Neural Sciences Division
Office of Naval Research, under
Contract No. N00014-88-K-0304
R&T 4421552

Robert J. Mislevy, Principal Investigator

Educational Testing Service
Princeton, New Jersey

October 1989

89 11 ... ...

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a REPORT SECURITY CLASSIFICATION: Unclassified	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S): RR-89-51-ONR	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION: Educational Testing Service; 6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION: Cognitive Science Program, Office of Naval Research (Code 1142CS), 800 North Quincy Street
6c ADDRESS (City, State, and ZIP Code): Princeton, NJ 08541	7b ADDRESS (City, State, and ZIP Code): Arlington, VA 22217-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION; 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-88-K-0304
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
61153N	RR04204	RR04204-01	R&T4421552

11 TITLE (Include Security Classification)
Integrating Cognitive and Psychometric Models to Measure Document Literacy (Unclassified)

12 PERSONAL AUTHOR(S)
Kathleen Sheehan and Robert J. Mislevy

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Technical	FROM ____ TO ____	October 1989	33

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Bayesian estimation; cognitive processing models; Item Response Theory; Linear Logistic Test Model; literacy assessment; National Assessment of Educational Progress
05	10		

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The Survey of Young Adult Literacy conducted in 1985 by the National Assessment of Educational Progress included sixty-three items that elicited skills in acquiring and using information from written documents. These items were analyzed in two distinct ways: (1) with an item response theory (IRT) model, which characterized items' difficulties and respondents' proficiencies as revealed simply by tendencies toward correct response; and (2) a qualitative cognitive model, which characterized items in terms of the processing tasks they required. This paper demonstrates how a generalization of Fischer and Scheiblechner's Linear Logistic Test Model can be used to integrate information from the cognitive analysis into the IRT analysis.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION: Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL: Dr. Charles E. Davis	22b TELEPHONE (Include Area Code): 202-696-4046; 22c OFFICE SYMBOL: ONR 1142CS

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603 Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

DD Form 1473, JUN 86 (Reverse)

Integrating Cognitive and Psychometric Models to Measure Document Literacy

Kathleen Sheehan

and

Robert J. Mislevy

Educational Testing Service

October 1989

This work was supported, in part, by Contract No. N00014-88-K-0304, R&T 4421552 from the Cognitive Sciences Program, Cognitive and Neural Sciences Division, Office of Naval Research. It does not necessarily reflect the views of that agency. We are grateful to Irwin Kirsch for his insights, his comments, and his patient explanations of the cognitive aspects of document literacy skills, and to Isaac Bejar and two anonymous reviewers for a number of helpful suggestions.

Integrating Cognitive and Psychometric Models

to Measure Document Literacy

Abstract

The Survey of Young Adult Literacy conducted in 1985 by the National Assessment of Educational Progress included sixty-three items that elicited skills in acquiring and using information from written documents. These items were analyzed in two distinct ways: (1) with an item response theory (IRT) model, which characterized items' difficulties and respondents' proficiencies as revealed simply by tendencies toward correct res[illegible] (2) a qualitative cognitive model, which characterize[illegible] s of the processing tasks they required. This p[illegible] es how a generalization of Fischer and Scheiblechner'[illegible] ear Logistic Test Model can be used to integrate information from the cognitive analysis into the IRT analysis.

Subject Terms: Bayesian estimation; cognitive processing models; Item Response Theory; Linear Logistic Test Model; literacy assessment; *National Assessment of Educational Progress*

1.0 Introduction

Perhaps the most important thrust in educational measurement today is, in Burstein's (1983) words, "... linking achievement testing to the cognitive processes employed in giving test responses and to the instructional experiences of students." Standard item-response theory and classical true-score psychometric models, while often providing practically useful summaries of the overall proficiencies of examinees and of the relative difficulties of items, do not do this. Cognitive-processing models, on the other hand, are typically qualitative, descriptive, and poorly suited to the broadly cast decision-making problems often encountered in educational practice. A recent line of development, therefore, has been to study the characteristics of psychometric items as cognitive tasks, using psychometric theory to summarize test data for action but cognitive theory to construct and analyze the test (Embretson, 1985).

This paper describes the implementation of such an approach in the construction and analysis of the Document Literacy scale in the Survey of Adult Literacy (Kirsch and Jungeblut, 1986), a study carried out under the auspices of the National Assessment of Educational Progress. After a brief overview of the Adult Literacy project, we outline (i) a cognitive-processing model proposed for solving the exercises, (ii) a psychometric model for the test, and (iii) a structure relating item parameters in the psychometric model to item features that are salient in the cognitive model, based on Mislevy's (1988) extension of

Scheiblechner (1972) and Fischer's (1973) linear logistic test model (LLTM).

2.0 An Overview of the NAEP Literacy Assessment

In 1984, the U.S. Department of Education provided funding for a nationwide assessment of the literacy skills of America's young adults, ages 21 through 25. The assessment was designed and carried out by the National Assessment of Educational Progress (NAEP) over the three year period from 1984 to 1986. A major innovation of the NAEP design was to call for a set of literacy tasks that simulate the diverse literacy demands of adult interactions in occupational, social, and educational settings. Implementation of this design led to a definition of literacy that encompassed three distinct skill areas:

o document literacy -- the skills needed to locate and use information contained in non-prose formats such as forms, tables, charts, signs/labels, indexes, schematics, and catalogues;

o prose literacy -- the skills needed to understand and use information from texts such as editorials, news stories and poems; and

o quantitative literacy -- the skills needed to perform arithmetic operations that are embedded in printed materials such as check book registers, order forms, and loan advertisements.

NAEP developed a total of ninety-three literacy tasks, sixty-three of which were classified as measuring document literacy, fifteen as measuring prose literacy, and fifteen as

measuring quantitative literacy. Most involved open-ended responses. For example, respondents were directed to: fill in a deposit slip; determine eligibility from a table of employee benefits; fill out an order form taken from a catalogue; and follow a set of directions to travel from one location to another using a map.

Trained interviewers administered the literacy tasks to a nationally representative household sample of approximately 3,600 young adults living in the 48 contiguous United States, using an item sampling design under which each task was administered to approximately 1,500 respondents. The procedures and the results of the assessment are detailed in Kirsch & Jungeblut (1986). In this paper, we describe a secondary analysis that was conducted to investigate correlates of task difficulty. Due to the small numbers of tasks available for measuring prose literacy and quantitative literacy, our analysis is restricted to the sixty-three tasks which comprise the document literacy scale.

3.0 A Cognitive Model for Document Literacy

A cognitive processing model for performance on document literacy tasks has been proposed by Kirsch and Mosenthal (1988). The model posits a solution process that can be summarized in the following four steps: (1) Identify the information given and requested in the task directive; (2) search the document until the requested information has been located; (3) make a match between the information identified in the document and the information

requested in the directive; and (4) determine whether the match adequately meets the criterion of the task.

As part of an earlier study of the factors influencing document task difficulty, Kirsch and Mosenthal developed a system to describe the complexity and organizational structure of documents and of the directives associated with document literacy tasks. This system, based on a significant revision of Mosenthal's (1985) taxonomic grammar of the expository continuum, characterizes the information contained in documents and document task directives according to three basic levels of organization: (1) the organizing category or OC, (2) the specific category or SPE, and (3) the semantic feature. These three levels of organization constitute three nested categories: semantic features are properties of pieces of information that belong to specific categories, which are nested within distinct organizing categories. Specific categories can also be nested within other specific categories. In fact, the more complex the document, the more likely it will be to find several levels of nesting of SPEs.

To illustrate these levels, consider the medicine label given in Figure 1. This document has three organizing categories: (1) the purpose for taking the medicine, (2) the recommended dosage levels, and (3) the list of cautions. Within the "Purpose" OC are two SPEs, one specifying that the medicine can be taken for "stuffed noses" and one specifying that it can also be taken for "running noses". The "Dosage" OC also contains two SPEs, one containing information specific to adult dosages and one

containing informati[illegible]pecific to children's dosages. The "Caution" OC, which is the most complex, contains four level-one SPE's and three level-two SPEs. These levels are illustrated in Figure 2, which provides a full linguistic representation, or parsing, of the medicine label. The reader should see Kirsch and Mosenthal (1988) for more information about this new grammar.

==================================

Insert Figures 1 and 2 about here

==================================

Based on this grammar, Kirsch and Mosenthal defined a number of variables, which, according to the processing model, would be expected to correlate with task difficulty. These variables have been classified into three distinct types: (1) Materials variables, which characterize the length and organizational complexity of the document to which a task refers; (2) Directive variables, which characterize the length and organizational complexity of the task directive; and (3) Process variables, which characterize the difficulty of the task solution process.

The Materials variables are

(1) the number of OCs in the document;

(2) the number of OCs in the document that are embedded;

(3) the deepest level of embedding for an OC;

(4) the number of SPEs in the document;

(5) the number of SPEs in the document that are embedded; and

(6) the deepest level of embedding for an SPE.

The Directive variables are

(1) the number of OCs in the directive;

(2) the number of OCs in the directive that are embedded;

(3) the deepest level of embedding for an OC;

(4) the number of SPEs in the Directive;

(5) the number of SPEs in the Directive that are Embedded; and

(6) the deepest level of embedding for an SPE.

The Process variables are defined as follows:

(1) Degree of Correspondence (DEGCORR). This variable refers to the explicitness of the match between the information requested in the directive or question and the corresponding information in the text. It is scored on an integer scale ranging from one to five with higher values indicating less explicit correspondence and therefore, more difficulty. For example, tasks requiring a single literal match are scored one, tasks requiring an inferential text-based match are scored three, and tasks requiring matches based on specialized prior knowledge are scor d five.

(2) Type of Information (TYPINFO). This variable concerns the type and number of restrictive conditions that must be held in mind in identifying and matching features. It too is scored on a one to five scale with lower values indicating less restrictive conditions.

(3) Plausibility of Distractors (DEGPLAUS). Document tasks typically require the examinee to skim an entire document in order to locate a piece of requested information. Since any piece of information embedded in the document could be interpreted as the requested information, the typical interpretation of the term "distractor", that is, the incorrect alternatives given with a

multiple-choice item, is not appropriate for document tasks. Instead, document task "distractors" include all pieces of information embedded in the document. The degree of plausibility of a distractor is measured by the extent to which the information embedded in the document shares semantic information with the correct answer to the question or directive, but does not satisfy all conditions specified. This variable is scored on a one to five scale with lower numbers indicating more shared semantic information and higher numbers indicating less.

The relationship between these three sets of variables and the four-step processing model can be stated as follows: The Directive variables characterize the difficulty of Step 1, identifying the information given and requested in the task directive; the Materials variables characterize the difficulty of Step 2, searching the document for requested information; and the Process variables characterize the difficulty of Steps 3 and 4, matching information and determining whether the criterion of the task has been satisfied.

Kirsch and Mosenthal (1988) succeeded in parsing sixty-one of the sixty-three document tasks, then scored the sixty-one in terms of the Materials, Directives, and Process variables using the scoring instructions in the appendices of their report. The results appear in Table 1; correlations among the variables appear in Table 2. (Because the level of OC and SPE embeddings for the document literacy task directives were almost entirely at the first level, not all of the directive embedding variables were

tabulated.) Task 46 is based on the Medicine Label. The reliability of the scoring was checked by training a third scorer and observing the proportion of exact agreement in rescores of one-third of the documents; the (very satisfactory) results are given in Table 3.

Tables 1-3 about here

Kirsch and Mosenthal regressed task proportions-correct on these task features in the total survey sample and in selected subpopulations. An adjusted R^2 of .87 resulted, with the strongest predictors being numbers and embedding of OCs, and the plausibility of distractors. This result provided empirical confirmation that the task attributes identified by the processing model did indeed largely account for task difficulty. The analysis addresses only average difficulty within populations, however, and provides no link between individuals' overall performance on the set of tasks and their expected success with documents and tasks with varying structures--the type of information required to target instruction to individual students and to design documents for specified types of users.

4.0 A Psychometric Model for Measuring Task Difficulty

In contrast, the expected outcomes of the confrontations between particular examinees and tasks are addressed by the response scaling methodology called item response theory (IRT; Lord, 1980). Unidimensional IRT models express the probability that an examinee will respond correctly to a particular test item

as a function of a single parameter that characterizes the proficiency of the examinee, and one or more additional parameters for each item that characterize measurement properties such as its difficulty. An important feature of IRT scaling is that the proficiency levels of all respondents can be reported on the same scale even when different individuals have been administered different subsets of tasks, as in the NAEP literacy assessment.

In this paper, we use the Rasch IRT model (Rasch, 1960) to exemplify the process of measuring task difficulty with a psychometric model. Let x_{ij} denote the response of examinee i to task j. Assume that responses are dichotomously scored, with 1 indicating a correct response and 0 indicating an incorrect response. The standard Rasch model gives the probability of a correct response as

$$\begin{aligned} P_j(\theta_i) &= P(x_{ij} = 1 | \theta_i, \beta_j) \\ &= \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} \end{aligned} \qquad (1)$$

where β_j characterizes the difficulty of task j and θ_i characterizes the proficiency of examinee i. Under the usual assumption of conditional independence, the probability of a respondent's pattern $\mathbf{x}_i = (x_{i1}, \ldots, x_{in})'$ of responses to n tasks is obtained as

$$P(\mathbf{x}_i | \theta_i, \boldsymbol{\beta}) = \prod_j P_j(\theta_i)^{x_{ij}} Q_j(\theta_i)^{1-x_{ij}} , \qquad (2)$$

where $Q_j(\theta) = 1 - P_j(\theta)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)'$. The probability of a data matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)'$ of responses from N examinees responding independently can be obtained as

$$P(\mathbf{X}|\boldsymbol{\theta},\boldsymbol{\beta}) = \prod_i P(\mathbf{x}_i|\theta_i,\boldsymbol{\beta}) \quad , \tag{3}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)'$. Once $\mathbf{X}$ has been observed, Equation 3 can be interpreted as a likelihood function, and provides a basis for estimating the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$.

Table 4 gives Rasch item parameter estimates obtained with Mislevy and Bock's (1984) BILOG computer program for the sixty-one literacy tasks that were parsed, on a scale in which the distribution of θ has a mean of zero and a standard deviation of one. Shown with estimates of the difficulty parameters are their (approximated) standard errors of estimation, or σ. Item 46 is the Medicine Label item, which with a difficulty parameter estimate of -2 is one of the easier items. A value of θ could be estimated for any respondent, and, via (1), the expectation of a correct response from that respondent to this item or any other could be calculated.

=================
Table 4 about here
=================

IRT models such as the Rasch model are widely accepted as useful tools for creating and analysing tests, adding precision and flexibility to the ways that examinees' proficiencies can be measured and compared. Note, however, that these models make no reference to the cognitive processes which an examinee must employ

in order to have a high probability of making a correct response; nor do they address the features of tasks that make them difficult. The model parameters merely indicate the relative proficiencies of respondents (θ) and the relative difficulties of tasks ($\boldsymbol{\beta}$) in the skill area considered.

5.0 An Integrated Approach

In a pioneering step toward integrating cognitive and psychometric models, Scheiblechner (1972) and Fischer (1973) posited a constrained Rasch model for item responses, the Linear Logistic Test Model (LLTM). In this model, task difficulty parameters are estimated as linear combinations of a smaller number of more elementary components. The elementary components are defined to reflect differences in the cognitive processing demands of the tasks. This approach represents a significant advance beyond standard IRT procedures, because it exploits auxilliary information about the cognitive processing demands of tasks to address <u>why</u> some tasks are more difficult than others.

To apply the LLTM to a set of test data, the usual response matrix $\mathbf{X}$ must be augmented with information pertaining to the processing demands of each test item. This information is expressed in terms of a set of K variables characterizing features of the items which are salient in the cognitive processing model. Examples include (i) Fischer's (1973) calculus example, in which items are characterized in terms of the number and type of operations a pupil must carry out in order to solve a differentiation problem, and (ii) the document literacy variables

which were defined in the previous section. Let $q_{1j},\ldots,q_{Kj}$ denote the item feature variables defined for the jth item. The LLTM assumes a Rasch model for task difficulty, but constrains the difficulty parameters β_j as follows:

$$\beta_j = \sum_{k=1}^{K} q_{kj}\, \eta_k \quad \text{for } j = 1,\ldots n\ , \tag{4}$$

or, in matrix notation $\boldsymbol{\beta} = Q'\boldsymbol{\eta}$, where Q' is an n by K matrix of item feature data and $\boldsymbol{\eta} = (\eta_1,\ldots,\eta_K)'$.

The original goal of explaining all of the reliable variation in item parameters by item features was not realized (Fischer and Formann, 1982), as rigorous tests of the sufficiency of the LLTM against the unconstrained model failed with few exceptions. It was often possible, however, to account for large portions of variation among item difficulties in terms of substantively meaningful item features, thus providing insights into the effects of educational treatments and helping to identify flawed items as unexpectedly easy or hard in light of the features that were expected to determine their operating characteristics.

A less restrictive method for incorporating cognitive processing information into a psychometric model has been proposed by Mislevy (1988). This alternative approach combines key aspects of the LLTM with the exchangeability concept of Bayesian inference (Lindley & Novick, 1981). As in the LLTM, differences in the cognitive processing demands of tasks are accounted for by regressing task difficulty on a smaller set of more elementary

components. Unlike the LLTM, however, parameter estimates obtained from the fitted regression model are not expected to account for <u>all</u> of the variation in true task difficulties. Instead, the expectation that true task difficulties will be distributed about the central values predicted by the fitted regression model is accounted for by (i) positing that the difficulty parameters of tasks with similar values of the item feature variables are exchangeable members of a common population; and (ii) imposing this task-population structure on the task difficulties, by means of Bayesian prior distributions.

In Mislevy's (1988) implementation of the approach, the prior distribution for individual task difficulties was assumed to be multivariate normal with mean $Q'\boldsymbol{\eta}$ and variance $\phi^2 I$, where the mean structure is defined as in the LLTM. This model was fitted as a two-stage empirical Bayes (EB) regression model: unconstrained difficulty parameters for individual tasks (as in Table 4), estimated in the first stage, provide data from which to estimate the unknown parameters η and ϕ^2 of the assumed item-parameter distribution in the second stage. Computational details are provided in that reference. Final task difficulty estimates $\bar{\beta}_j$ are precision-weighted combinations of the unrestricted Rasch estimates $\hat{\beta}_j$ and the regression estimates $q_j'\boldsymbol{\eta}$:

$$\bar{\beta}_j = (w_{1j} q_j' \boldsymbol{\eta} + w_{2j}\hat{\beta}_j)/(w_{1j} + w_{2j})$$

where $w_{1j} = 1/\hat{\phi}^2$ and $w_{2j} = 1/\hat{\sigma}^2_j$. The final task difficulty estimates can be viewed as a compromise between LLTM estimates, where items with identical features are constrained to have identical difficulty estimates, and standard Rasch difficulty estimates, where information about item features is ignored.

Like the LLTM, this approach provides a link between the cognitive processing model assumed to be influencing task responses and the tasks' resulting difficulties. To the extent that the structural model for item parameters fits, it provides a basis for understanding just what makes items difficult. It is a powerful argument for the construct validity of a test if it can be shown that item difficulties are determined predominantly by manipulable features in a cognitive model built around the skills intended to be measured (Embretson, 1985). To the extent that the model does <u>not</u> fit, it identifies unexpectedly hard or easy items, information that should prove useful for item construction.

6.0 Application to the Document Literacy Scale

As described above, both the cognitive processing analysis and the psychometric analysis were first applied to the Document Literacy data separately. The variables in Table 1, resulting from parsing the tasks, signify salient features of the items as indicated by the cognitive processing model, and provide insights into their processing requirements. The unrestricted Rasch difficulty estimates ($\hat{\beta}$) in Table 4 indicate the difficulty of tasks from a purely empirical point of view. We now apply the integrated model described in the preceeding section.

In considering variables to include in the augmented data matrix, Kirsch and Mosenthal's (1988) results were used to eliminate three of the parsing variables: (i) the deepest level of OC embedding in the Materials, (ii) the deepest level of SPE embedding in the Materials, and (iii) the deepest level of OC embedding in the Directives. Univariate distributions were tabulated for the nine remaining item feature variables, and transformations were applied to eliminate extreme asymmetries: a square root transformation for the "Number of OC's" variable, a logarithmic transformation for "Number of SPE's", and logit transformations for "Number of Embedded OC's" and "Number of Embedded SPE's" after expressing them as proportions of total OC's and SPE's respectively. In addition, both the Materials variables and the Directive variables were centered and scaled to have a mean of zero and variance 1. Because the Process variables represent ordered categories, rather than counts, these variables were centered by recoding the original values of 1 to 5 as -1 to 3. These rescaled variables were used in all subsequent analyses.

The parameter estimates obtained from fitting a two-stage Empirical Bayes regression model to these data are given in Table 5. They include the estimated coefficients for the intercept term and the nine item feature variables $(\hat{\eta}_0, \hat{\eta}_1, \ldots, \hat{\eta}_9)$, and the estimated standard deviation $\hat{\phi}$ for the normal distribution of residuals of the task difficulty parameters from their expected values. Because the model was estimated from standardized data,

the magnitude of the coefficients provide an indication of the relative contribution of each variable to expected difficulty.

Insert Table 5 about here

To further investigate the contribution of each item feature variable to variation in predicted task difficulty, three alternative models were estimated: (1) a model that excluded the Materials variables; (2) a model that excluded the Directive variables; and (3) a model that excluded the Process variables. The estimated coefficients for these three alternative models are also shown in Table 5. Note the similarity of the coefficients listed for the Materials variables in the Full model and in the model which excluded the Directive variables (Model #2), and the similarity of the coefficients listed for the Directive variables in the Full model and in the model which excluded the Materials variables (Model #1). These similarities are a result of the low correlation between the Materials variables and the Directive variables. By contrast, the coefficients of both the Materials variables and the Directive variables changed from the Full model to the model which excluded the Process variables (Model #3). These changes are a result of the higher correlations between the Process variables and the Materials variables and between the Process variables and the Directive variables. Because Model #3 is not contaminated by Process variable correlation, its coefficients provide the most accurate picture of the relative contributions to predicted task difficulty provided by the

Materials variables and the Directive variables. In particular, when the process variables are excluded, task difficulty increases most rapidly with the No. of SPEs in the Materials and the No. of SPEs in the Directive. Increasing the No. of OC's in the Directive and in the Materials also increases task difficulty, but not by as much. By far, the smallest contribution to task difficulty is provided by the OC and SPE embedding variables.

Table 5 also lists approximate R^2 values for each model. In the standard regression setting, the R^2 statistic is calculated as the ratio of explained variation to total variation. In this application, true task difficulties are unobservable so total variation is approximated using the variation observed in the EB estimates $\bar{\beta}$. Several conclusions can be drawn from the R^2 values. First, differences in the cognitive processing demands of document literacy tasks, as measured by the cognitive processing variables proposed by Kirsch and Mosenthal, account for approximately 80% of the observed variation in task difficulty. Second, the largest contribution to explained variation is provided by the Process variables. When these variables were excluded from the model, the R^2 statistic dropped by more than 20 points. This indicates that the Process variables are tapping an aspect of task difficulty that is not well predicted by either the Materials variables or the Directive variables. Third, the five point decreases in the R^2 values listed for Alternative Models #1 and #2 indicate that both the Materials variables and the Directive variables are also measuring unique aspects of task difficulty. Thus, although the

Process variables appear to be the most important, neither the Materials variables nor the Directive variables, can be excluded without diminishing predictive capability.

Figure 3 plots the residuals obtained from fitting the full model against percent correct. Negative residuals indicate that the task was easier than predicted, that is, easier than other tasks with similar values of the item feature variables. The plot shows a scatter of low positive and negative residuals among tasks with percent correct values below 90 percent. This suggests that the item feature variables have been successful at predicting task difficulty among tasks with low percent correct values. However, several high negative residuals occur among the tasks with percent correct values above 90 percent. This suggests that the item feature variables have not provided useful information pertaining to gradations of difficulty among extremely easy tasks.[1]

========================

Insert Figure 3 about here

========================

7.0 Discussion

The two-stage Empirical Bayes regression model provides a link between Kirsch and Mosenthal's cognitive model for solving document literacy tasks and the psychometric IRT model for task difficulty. The integrated approach led to the following findings: (i) document literacy task difficulty was highly related

[1]This explains why the R^2 is slightly lower in this analysis than in Kirsch and Mosenthal's regression analysis of percents-correct: task features account poorly for differences among easy items, which are minimized in the percent-correct metric but expanded in the Rasch difficulty (logit) metric.

to the Process variables and somewhat less related to the Materials variables and the Directive variables; and (ii) the cognitive model for explaining task difficulty was deficient at explaining gradations of difficulty among extremely easy tasks. Of course these results are based on only the present data, which effectively fit a regression model with nine independent variables to sixty-one observations. Extensions of the literacy survey currently in progress, however, should yield response data on as many as a hundred new document literacy tasks written to similar specifications. If these subsequent assessments reveal similar findings, an examination of tasks with high negative residuals will be conducted in order to determine factors associated with extremely easy document literacy tasks. Knowledge of such factors should prove useful for document design and construction

It is increasingly becoming recognized that merely high reliability coefficients do not guarantee a "good" test, nor do high predictive relationships guarantee a "valid" one. The onus has been placed (appropriately!) upon the tester to demonstrate that the skills tapped in an educational test are in fact those deemed important to measure. The two-stage approach exemplified in this paper capitalizes upon advances in the psychometric and cognitive disciplines to address this need. IRT models, which provide measures of overall proficiency for making decisions about individual examinees, also define implicitly the variable being measured through implications of correct response at the various levels of proficiency. A demonstration that this empirical

characterization of proficiency can be largely accounted for by the key features of items from the perspective of a cognitive model argues strongly for the construct validity of the measure, constitutes a theoretical foundation for further item development, and provides an additional means of detecting items that tap irrelevant skills.

References

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM-algorithm. Psychometrika, 46, 443-459.

Burstein, L. (1983). A word about this issue (editor's note). Journal of Educational Measurement, 20, 99-102.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion. Journal of the Royal Statistical Society, Series B, 39, 1-38.

Embretson, S.E. (Ed.) (1985). Test Design: Developments in psychology and psychometrics. Orlando, FL: Academic Press.

Fischer G.H. (1973). The linear logistic test model as an instrument in educational research. Acta Psychologica, 37, 359-374.

Fischer, G.H., and Formann, A.K. (1982). Some applications of logistic latent trait models with linear constraints on the parameters. Applied Psychological Measurement, 6, 397-416

Kirsch, I.S., and Jungeblut, A. (1986). Literacy: Profiles of America's young adults (Final Report 16-PL-01). Princeton, NJ: Educational Testing Service.

Kirsch, I.S., and Mosenthal, P.B. (1988). Understanding document literacy: Variables underlying the performance of young adults (Research Report RR-88-62). Princeton, NJ: Educational Testing Service.

Lindley, D.V., and Novick, M.R. (1981). The role of exchangeability in inference. Annals of Statistics, 9, 45-58.

Mislevy, R.J. (1988). Exploiting collateral information about items in the estimation of Rasch item difficulty parameters. Applied Psychological Measurement, 12, 281-296.

Mislevy, R.J., and Bock, R.D. (1984). BILOG: Item analysis and test scoring with binary logistic models [Computer program]. Mooresville, IN: Scientific Software.

Mosenthal, P.B. (1985). Defining the expository discourse continuum: Towards a taxonomy of expository text types. Poetics, 14, 387-414.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research.

Scheiblechner, H. (1972). Das lernen und losen komplexer denkaufgaben. Zeitschrift fur experimentalle und Angewandte Psychologie, 19, 476-506.

Table 1

Cognitive Processing Variables
for 61 Document Literacy Tasks

	Materials						Directives			Process		
Task	No. OCs	No. Em.OCs	Deep. Em.OC	No. SPEs	No. Em.SPE	Deep. Em.SPE	No. OCs	Deep. Em.OC	No. SPEs	TYP INFO	DEG PLAUS	DEG CORR
1	14	2	2	17	1	2	1	1	1	1	2	1
2	1	0	1	7	0	1	1	1	1	1	1	1
3	14	0	1	14	1	2	1	1	1	1	2	2
4	14	0	1	14	1	2	1	1	1	1	1	1
5	17	1	2	20	1	2	1	1	1	1	1	1
6	6	0	1	6	1	2	1	1	1	1	2	2
7	17	1	2	20	1	2	1	1	1	1	1	1
8	17	1	2	20	1	2	1	1	1	1	1	1
9	36	32	3	19	0	1	1	1	1	1	1	1
10	14	0	1	14	1	2	1	1	1	1	2	1
11	14	0	1	25	2	3	1	1	1	2	2	2
12	14	0	1	25	2	3	1	1	2	3	2	2
13	14	0	1	14	1	2	1	2	1	1	2	2
14	14	0	1	14	1	2	1	2	1	1	2	1
15	17	12	2	13	0	1	3	3	1	4	2	2
16	14	0	1	14	1	2	1	1	3	4	3	2
17	14	0	1	25	2	3	2	1	5	4	4	2
18	36	32	3	19	0	1	2	2	1	1	3	1
19	34	30	3	329	3	2	2	2	1	4	5	1
20	34	30	3	329	3	2	1	1	2	5	5	1
21	34	30	3	329	3	2	3	3	3	5	3	1
22	34	30	3	329	3	2	3	3	2	3	3	1
23	14	7	3	65	33	5	1	1	2	1	2	1
24	8	1	2	39	0	1	1	1	2	1	2	1
25	10	2	2	83	15	3	1	1	1	1	1	1
26	10	2	2	83	15	3	4	1	4	3	2	1
27	8	4	2	85	38	3	2	1	2	1	2	1
28	5	0	1	64	0	1	2	1	2	1	2	1
29	5	0	1	64	0	1	2	1	2	3	2	1
30	16	3	2	30	2	2	3	1	3	1	3	1

Table 1 (Continued)

Cognitive Processing Variables
for 61 Document Literacy Tasks

	Materials						Directives			Process		
Task	No. OCs	No. Em.OCs	Deep. Em.OC	No. SPEs	No. Em.SPE	Deep. Em.SPE	No. OCs	Deep. Em.OC	No. SPEs	TYP INFO	DEG PLAUS	DEG COR
31	16	3	2	80	2	2	2	1	2	1	3	1
32	2	0	1	11	3	2	1	1	1	1	2	1
33	4	0	1	16	3	3	0	0	2	1	2	1
34	13	0	1	102	5	2	1	1	1	3	2	2
35	16	5	2	990	0	1	3	1	3	4	3	1
36	0	0	0	51	8	2	0	0	2	4	2	1
37	61	55	2	188	2	2	1	1	1	1	2	1
38	11	0	1	16	3	3	1	1	1	1	1	1
39	11	0	1	16	3	3	2	1	2	1	2	1
40	11	0	1	16	3	3	1	1	1	1	2	1
41	37	0	1	173	10	2	1	1	1	1	2	1
42	22	3	2	58	9	3	1	1	1	1	1	2
43	22	3	2	58	9	3	1	1	2	1	1	1
44	22	3	2	58	9	3	1	1	1	1	1	1
45	22	3	2	58	9	3	1	1	1	1	1	2
46	3	0	1	37	34	2	1	1	1	1	2	1
47	36	0	1	225	33	2	1	1	1	4	3	1
48	36	0	1	225	33	2	1	1	3	2	3	1
49	36	0	1	225	33	2	1	1	2	2	2	1
50	1	0	1	29	16	3	0	0	2	2	3	1
51	18	16	2	135	0	1	1	1	2	1	2	1
52	6	0	1	65	0	1	1	0	2	1	2	1
53	108	0	1	234	36	1	1	1	1	3	2	1
54	18	0	1	574	3	2	1	1	1	1	2	1
55	64	0	1	716	57	2	1	1	1	3	2	1
56	135	0	1	584	0	1	1	1	2	5	3	4
57	21	0	1	438	0	1	4	1	3	4	3	1
58	103	6	1	379	24	2	1	1	1	3	2	4
59	38	30	3	270	4	3	3	1	2	5	2	5
60	38	30	3	270	4	3	2	1	1	5	3	1
61	7	0	1	11	0	1	1	1	1	1	1	1

Table 2

Intercorrelations among Item Features

	Materials						Directives			Process		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<u>Materials</u>												
(1) No. of OCs	1.00	.25	.09	.52	.31	-.20	-.00	.13	-.10	-.45	-.40	-.19
(2) No. of OCs Embedded		1.00	.74	.18	-.18	-.05	.29	.41	-.04	-.02	-.29	-.34
(3) Levels of OC Embeddings			1.00	.12	-.12	.15	.41	.44	.03	.03	-.16	-.21
(4) No. of SPEs				1.00	.25	-.23	.31	.11	.17	-.15	-.53	-.39
(5) No. of SPEs Embedded					1.00	.26	-.15	-.13	-.02	.08	-.05	.00
(6) Levels of SPE Embeddings						1.00	-.13	-.17	.08	-.08	.09	.09
<u>Directives</u>												
(7) No. of OCCs							1.00	.50	.50	-.07	-.41	-.32
(8) Levels of OC Embeddings								1.00	-.06	-.03	-.22	-.21
(9) No. of SPEs									1.00	-.02	-.40	-.46
<u>Process</u>												
(10) Degrees of Correspondence										1.00	-.38	-.62
(11) Type of Information											1.00	-.03
(12) Plausibility of Distractors												1.00

Table 3

Proportions of Exact Agreement Among Raters

Variable	Proportion of Agreement
Materials Variables	
Number of OCs	100 %
Number of Embedded OCs	100 %
Level of OC Embedding	98 %
Number of SPEs	96 %
Number of Embedded SPEs	93 %
Level of SPE Embedding	88 %
Directive Variables	
Number of OCs	96 %
Level of OC Embedding	99 %
Number of SPEs	90 %
Process Variables	
Degrees of Correspondence	95 %
Type of Information	86 %
Plausibility of Distractors	90 %

Table 4

Results of Fitting an Unrestricted Rasch Model

Task	$\hat{\beta}$	$\hat{\sigma}$	% Correct	Task	$\hat{\beta}$	$\hat{\sigma}$	% Correct
1	-4.051	0.120	99	31	-1.110	0.054	79
2	-3.503	0.088	98	32	-2.128	0.047	91
3	-3.277	0.126	97	33	-2.412	0.053	94
4	-3.198	0.121	97	34	-0.912	0.051	76
5	-3.468	0.147	96	35	-0.201	0.047	56
6	-2.638	0.058	96	36	-1.016	0.053	80
7	-4.153	0.218	96	37	-2.233	0.078	94
8	-2.914	0.110	94	38	-2.641	0.093	96
9	-2.758	0.098	94	39	-1.157	0.055	81
10	-1.967	0.070	91	40	-2.129	0.075	93
11	-1.590	0.060	89	41	-2.920	0.110	94
12	-1.104	0.053	81	42	-1.842	0.067	90
13	-2.247	0.078	92	43	-1.894	0.068	90
14	-1.252	0.056	80	44	-1.819	0.066	89
15	-1.217	0.057	80	45	-1.883	0.068	91
16	-0.420	0.048	68	46	-2.062	0.071	90
17	-0.384	0.046	68	47	-1.133	0.053	78
18	-1.802	0.066	88	48	-1.245	0.055	79
19	-0.613	0.048	69	49	-1.409	0.057	85
20	-0.203	0.046	62	50	-1.884	0.069	86
21	0.294	0.045	48	51	-2.413	0.083	94
22	-0.471	0.047	67	52	-1.783	0.066	89
23	-1.734	0.063	89	53	-1.365	0.057	84
24	-1.968	0.068	92	54	-1.622	0.062	37
25	-1.896	0.066	90	55	-1.095	0.054	81
26	-0.457	0.047	67	56	0.115	0.046	52
27	-1.712	0.063	88	57	-0.467	0.047	62
28	-1.860	0.066	88	58	-0.162	0.046	63
29	-0.749	0.049	73	59	1.244	0.053	28
30	-0.567	0.048	68	60	0.055	0.046	59
				61	-2.726	0.096	97

Note: Rasch difficulty estimates are not strictly monotonically related to proportions correct in this analysis because of the matrix-sampling data collection design; the percents-correct reflect performance in different randomly equivalent samples of respondents

Table 5

Estimated Regression Parameters
and Approximate R^2 Values

Variable	Type	Full Model	Alternative Models #1	#2	#3
Intercept		-1.404	-1.462	-1.409	-1.603
No.OCs	MAT	-0.096	e	-0.191	0.157
No.Emb.OCs	MAT	0.024	e	0.048	0.069
No.SPEs	MAT	0.383	e	0.442	0.459
No.Emb.SPEs	MAT	0.159	e	0.090	0.099
No.OCs	DIR	0.212	0.210	e	0.245
No.SPEs	DIR	0.149	0.163	e	0.364
TYPINFO	PROC	0.268	0.351	0.327	e
DEGPLAUS	PROC	0.202	0.229	0.264	e
DEGCORR	PROC	0.360	0.285	0.372	e
Std.Dev. (ϕ)		0.467	0.538	0.534	0.689
Approximate R^2		.81	.75	.76	.59

e=variable was intentionally excluded from the model

For Stuffed and Running Noses:

Dosage:
Adults - 2 teaspoons every 4 hours; Children over 6 years - 1 teaspoon every 4 hours.

Caution:
Unless directed by physician, do not exceed recommended dosage. If drowsiness occurs, do not drive or operate dangerous machinery. Individuals with high blood pressure, heart disease, diabetes, or thyroid disease should use only as directed by a physician.

Figure 1. The Medicine Label document.

```
      *|\OC purpose
       |      |\SPE For Stuffed Noses
       |  AND \SPE For Running Noses
  *AND |\OC Dosage
       |      |\SPE *take
       |      |      |\AG  Adults
       |      |      |\OBJ teaspoons
       |      |      |     \ATT 2
       |      |       \TEMP hours
       |      |            |\ATT 4
       |                    \ATT every
       |  AND \SPE *take
       |             |\AG children
       |             |      \ATT over six
       |             |\ATT  teaspoon
       |             |      \ATT 1
       |              \TEMP hours
       |                   |\ATT 4
       |                    \ATT every
  *AND \OC caution
            |\SPE do exceed
            |     *|\AG you
            |      |\OBJ dosage
            |      |      \ATT recommended
            |      |\NEG not
     unless | COND \SPE directed
            |           |\AGT by physician
            |          * \OBJ you
       *AND |\SPE do drive
         OR |\SPE do operate
            |    *|\AG you
            |     |\OBJ machinery
            |     |     \ATT dangerous
            |     |\NEG not
          If|COND\SPE occurs
            |          \AG drowsiness
       *AND  \SPE should use
                  |\AG individuals with
                  |   *OR|\ATT blood pressure
                  |      |       \ATT high
                  |   *OR|\ATT heart disease
                  |      |      *\ATT high
                  |   *OR|\ATT diabetes
                  |      |      *\ATT high
                  |    OR \ATT thyroid disease
                  |             *\ATT high
          as COND \SPE directed
                       |\MAN only
                        \AG by physician
```

Figure 2. A parsing of the Medicine Label document.

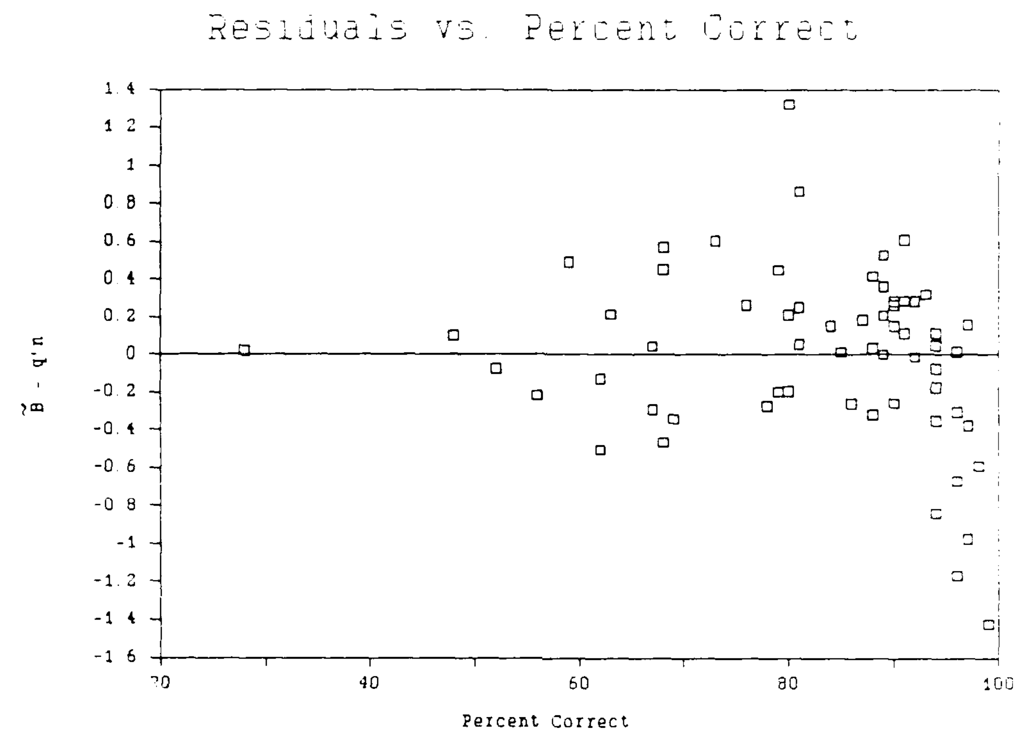

Figure 3. The full model residuals plotted against percent correct for 61 document literacy tasks.

Dr. Terry Ackerman
American College Testing Programs
P.O. Box 168
Iowa City, IA 52243

Dr. Robert Ahlers
Code N711
Human Factors Laboratory
Naval Training Systems Center
Orlando, FL 32813

Dr. James Algina
1403 Norman Hall
University of Florida
Gainesville, FL 32605

Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

Dr. Eva L. Baker
UCLA Center for the Study
of Evaluation
145 Moore Hall
University of California
Los Angeles, CA 90024

Dr. Isaac Bejar
Mail Stop: 10-R
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Ramat Aviv 69978
ISRAEL

Dr. Arthur S. Blaiwes
Code N712
Naval Training Systems Center
Orlando, FL 32813-7100

Dr. Bruce Bloxom
Defense Manpower Data Center
99 Pacific St.
Suite 155A
Monterey, CA 93943-3231

Dr. R. Darrell Bock
University of Chicago
NORC
6030 South Ellis
Chicago, IL 60637

Cdt. Arnold Bohrer
Sectie Psychologisch Onderzoek
Rekruterings-En Selectiecentrum
Kwartier Koningen Astrid
Bruijnstraat
1120 Brussels, BELGIUM

Dr. Robert Breaux
Code 7B
Naval Training Systems Center
Orlando, FL 32813-7100

Dr. Robert Brennan
American College Testing
Programs
P. O. Box 168
Iowa City, IA 52243

Dr. John B. Carroll
409 Elliott Rd., North
Chapel Hill, NC 27514

Dr. Robert M. Carroll
Chief of Naval Operations
OP-01B2
Washington, DC 20350

Dr. Raymond E. Christal
UES LAMP Science Advisor
AFHRL/MOEL
Brooks AFB, TX 78235

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
Los Angeles, CA 90089-1061

Director,
Manpower Support and
Readiness Program
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Dr. Stanley Collyer
Office of Naval Technology
Code 222
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Hans F. Crombag
Faculty of Law
University of Limburg
P.O. Box 616
Maastricht
The NETHERLANDS 6200 MD

Dr. Timothy Davey
American College Testing Program
P.O. Box 168
Iowa City, IA 52243

Dr. C. M. Dayton
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Ralph J. DeAyala
Measurement, Statistics,
and Evaluation
Benjamin Bldg., Rm. 4112
University of Maryland
College Park, MD 20742

Dr. Dattprasad Divgi
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Hei-Ki Dong
Bell Communications Research
6 Corporate Place
PYA-1K226
Piscataway, NJ 08854

Dr. Fritz Drasgow
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC
(12 Copies)

Dr. Stephen Dunbar
224B Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. James A. Earles
Air Force Human Resources Lab
Brooks AFB, TX 78235

Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence, KS 66045

Dr. George Englehard, Jr.
Division of Educational Studies
Emory University
210 Fishburne Bldg.
Atlanta, GA 30322

Dr. Benjamin A. Fairbank
Performance Metrics, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

Dr. P-A. Federico
Code 51
NPRDC
San Diego, CA 92152-6800

Dr. Leonard Feldt
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Richard L. Ferguson
American College Testing
P.O. Box 168
Iowa City, IA 52243

Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

Dr. Myron Fischl
U.S. Army Headquarters
DAPE-MRR
The Pentagon
Washington, DC 20310-0300

Prof. Donald Fitzgerald
University of New England
Department of Psychology
Armidale, New South Wales 2351
AUSTRALIA

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Alfred R. Fregly
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Robert D. Gibbons
Illinois State Psychiatric Inst.
Rm 529W
1601 W. Taylor Street
Chicago, IL 60612

Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01003

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

DORNIER GMBH
P.O. Box 1420
D-7990 Friedrichshafen 1
WEST GERMANY

Prof. Edward Haertel
School of Education
Stanford University
Stanford, CA 94305

Dr. Ronald K. Hambleton
University of Massachusetts
Laboratory of Psychometric
and Evaluative Research
Hills South, Room 152
Amherst, MA 01003

Dr. Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign, IL 61820

Dr. Grant Henning
Senior Research Scientist
Division of Measurement
Research and Services
Educational Testing Service
Princeton, NJ 08541

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 63
San Diego, CA 92152-6800

Dr. Paul W. Holland
Educational Testing Service, 21-T
Rosedale Road
Princeton, NJ 08541

Prof. Lutz F. Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 92010

Mr. Dick Hoshaw
OP-135
Arlington Annex
Room 2834
Washington, DC 20350

Dr. Lloyd Humphreys
University of Illinois
Department of Psycnology
603 East Daniel Street
Champaign, IL 61820

Dr. Steven Hunka
3-104 Educ. N.
University of Alberta
Edmonton, Alberta
CANADA T6G 2G5

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Robert Jannarone
Elec. and Computer Eng. Dept.
University of South Carolina
Columbia, SC 29208

Dr. Douglas H. Jones
Thatcher Jones Associates
P.O. Box 6640
10 Trafalgar Court
Lawrenceville, NJ 08648

Dr. Brian Junker
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Milton S. Katz
European Science Coordination Office
U.S. Army Research Institute
Box 65
FPO New York 09510-1500

Prof. John A. Keats
Department of Psychology
University of Newcastle
N.S.W. 2308
AUSTRALIA

Dr. G. Gage Kingsbury
Portland Public Schools
Research and Evaluation Department
501 North Dixon Street
P. O. Box 3107
Portland, OR 97209-3107

Dr. William Koch
Box 7246, Meas. and Eval. Ctr.
University of Texas-Austin
Austin, TX 78703

Dr. Leonard Kroeker
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Dr. Jerry Lehnus
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn, VA 22209

Dr. Thomas Leonard
University of Wisconsin
Department of Statistics
1210 West Dayton Street
Madison, WI 53705

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Charles Lewis
Educational Testing Service
Princeton, NJ 08541-0001

Dr. Robert L. Linn
Campus Box 249
University of Colorado
Boulder, CO 80309-0249

Dr. Robert Lockman
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

Dr. George B. Macready
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

Dr. James R. McBride
The Psychological Corporation
1250 Sixth Avenue
San Diego, CA 92101

Dr. Clarence C. McCormick
HQ, USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago, IL 60064

Dr. Robert McKinley
Law School Admission Services
Box 40
Newtown, PA 18940

Dr. James McMichael
Technical Director
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Dr. William Montague
NPRDC Code 13
San Diego, CA 92152-6800

Ms. Kathleen Moreno
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Headquarters Marine Corps
Code MPI-20
Washington, DC 20380

Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Norman, OK 73071

Deputy Technical Director
NPRDC Code 01A
San Diego, CA 92152-6800

Director, Training Laboratory,
NPRDC (Code 05)
San Diego, CA 92152-6800

Director, Manpower and Personnel
Laboratory,
NPRDC (Code 06)
San Diego, CA 92152-6800

Director, Human Factors
& Organizational Systems Lab,
NPRDC (Code 07)
San Diego, CA 92152-6800

Library, NPRDC
Code P201L
San Diego, CA 92152-6800

Commanding Officer,
Naval Research Laboratory
Code 2627
Washington, DC 20390

Dr. Harold F. O'Neil, Jr.
School of Education - WPH 801
Department of Educational
Psychology & Technology
University of Southern California
Los Angeles, CA 90089-0031

Dr. James B. Olsen
WICAT Systems
1875 South State Street
Orem, UT 84058

Office of Naval Research,
Code 1142CS
800 N. Quincy Street
Arlington, VA 22217-5000
(6 Copies)

Office of Naval Research,
Code 125
800 N. Quincy Street
Arlington, VA 22217-5000

Assistant for MPT Research,
Development and Studies
OP 01B7
Washington, DC 20370

Dr. Judith Orasanu
Basic Research Office
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

Dr. Peter J. Pashley
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036

Dr. James Paulson
Department of Psychology
Portland State University
P.O. Box 751
Portland, OR 97207

Dept. of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey, CA 93943-5026

Department of Operations Research,
Naval Postgraduate School
Monterey, CA 93940

Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Malcolm Ree
AFHRL/MOA
Brooks AFB, TX 78235

Mr. Steve Reiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455-0344

Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg.
Knoxville, TN 37916-0900

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152-6800

Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

Dr. Mary Schratz
905 Orchid Way
Carlsbad, CA 92009

Dr. Dan Segall
Navy Personnel R&D Center
San Diego, CA 92152

Dr. W. Steve Sellman
OASD(MRA&L)
2B269 The Pentagon
Washington, DC 20301

Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujisawa 251
JAPAN

Dr. William Sims
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. H. Wallace Sinaiko
Manpower Research
and Advisory Services
Smithsonian Institution
801 North Pitt Street, Suite 120
Alexandria, VA 22314-1713

Dr. Richard E. Snow
School of Education
Stanford University
Stanford, CA 94305

Dr. Richard C. Sorensen
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Judy Spray
ACT
P.O. Box 168
Iowa City, IA 52243

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. Peter Stoloff
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

Mr. Brad Sympson
Navy Personnel R&D Center
Code-131
San Diego, CA 92152-6800

Dr. John Tangney
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Kikumi Tatsuoka
CERL
252 Engineering Research
Laboratory
103 S. Mathews Avenue
Urbana, IL 61801

Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

Mr. Gary Thomasson
University of Illinois
Educational Psychology
Champaign, IL 61820

Dr. Robert Tsutakawa
University of Missouri
Department of Statistics
222 Math. Sciences Bldg.
Columbia, MO 65211

Dr. Ledyard Tucker
University of Illinois
Department of Psychology
603 E. Daniel Street
Champaign, IL 61820

Dr. David Vale
Assessment Systems Corp.
2233 University Avenue
Suite 440
St. Paul, MN 55114

Dr. Frank L. Vicino
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Howard Wainer
Educational Testing Service
Princeton, NJ 08541

Dr. Ming-Mei Wang
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Thomas A. Warm
FAA Academy AAC934D
P.O. Box 25082
Oklahoma City, OK 73125

Dr. Brian Waters
HumRRO
12908 Argyle Circle
Alexandria, VA 22314

Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455-0344

Dr. Ronald A. Weitzman
Box 146
Carmel, CA 93921

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Douglas Wetzel
Code 51
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Rand R. Wilcox
University of Southern
California
Department of Psychology
Los Angeles, CA 90089-1061

German Military Representative
ATTN: Wolfgang Wildgrube
Streitkraefteamt
D-5300 Bonn 2
4000 Brandywine Street, NW
Washington, DC 20016

Dr. Bruce Williams
Department of Educational
Psychology
University of Illinois
Urbana, IL 61801

Dr. Hilda Wing
NRC MH-176
2101 Constitution Ave.
Washington, DC 20418

Mr. John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. George Wong
Biostatistics Laboratory
Memorial Sloan-Kettering
Cancer Center
1275 York Avenue
New York, NY 10021

Dr. Wallace Wulfeck, III
Navy Personnel R&D Center
Code 51
San Diego, CA 92152-6800

Dr. Kentaro Yamamoto
03-T
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

Dr. Joseph L. Young
National Science Foundation
Room 320
1800 G Street, N.W.
Washington, DC 20550

Mr. Anthony R. Zara
National Council of State
Boards of Nursing, Inc.
625 North Michigan Avenue
Suite 1544
Chicago, IL 60611

Dr. Ratna Nandakumar
Dept. of Educational Studies
Willard Hall, Room 213
University of Deleware
Newark, DE 19716